

Геномные перестройки

Николай Вяххи

vyahhi@bioinformaticsinstitute.ru

Computer Science Club
Санкт-Петербург, 2013



**Институт
Биоинформатики**



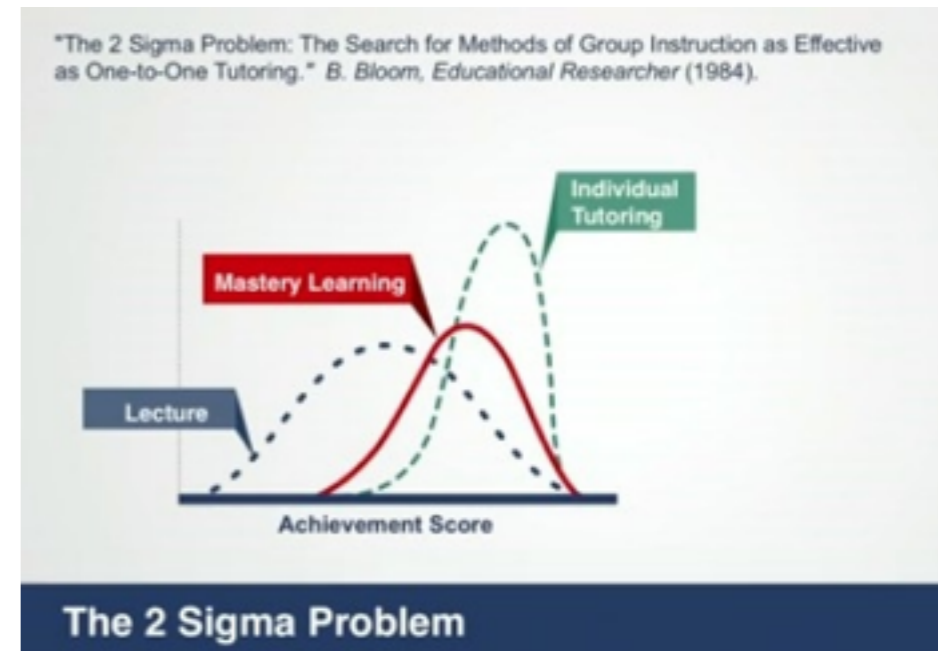
Формат обучения

12 лекций по воскресеньям

Квизы для самопроверки

Домашние задания и вопросы онлайн

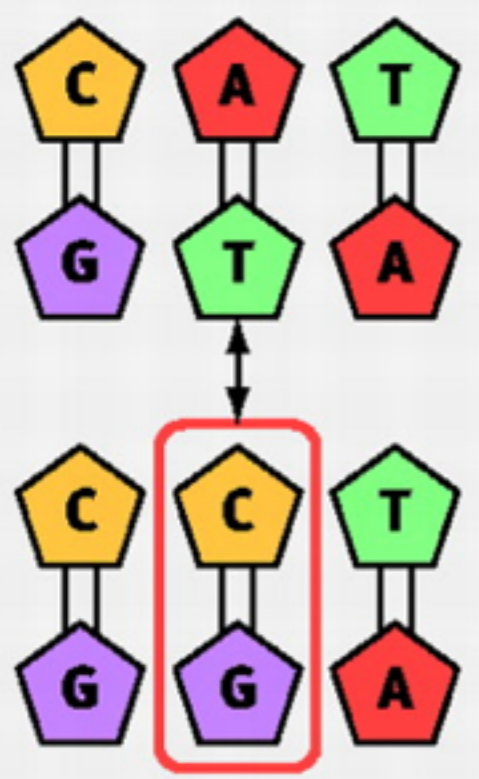
<http://rosalind.info/classes/enroll/ff45302de4/>



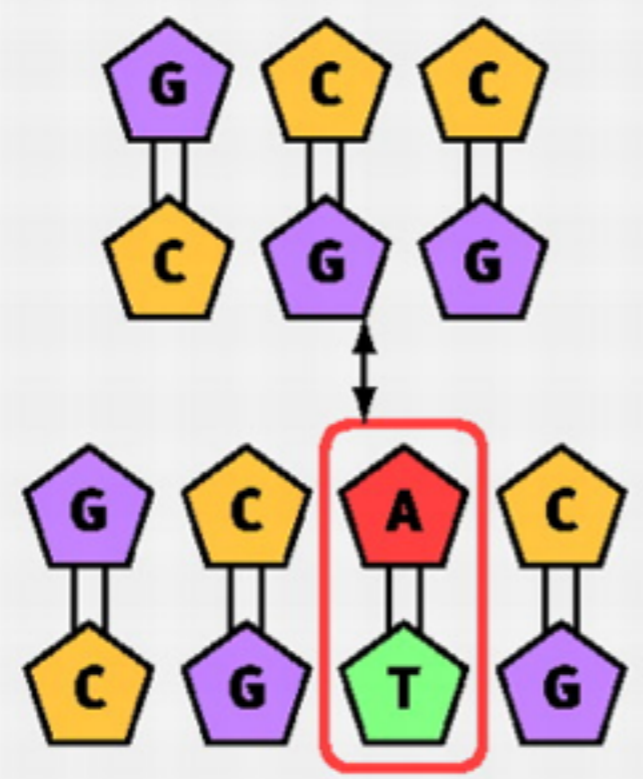
ОТВЕТЫ НА КВИЗ

1. ДНК → РНК → Белки
2. Замена, вставка, удаление символа
3. $210 = 20 \cdot 19 / 2 + 20$
4. Нахождение глобального выравнивания
5. Нахождение локального выравнивания
6. $\text{GapOpening} + (L-1) \cdot \text{GapExtension}$
7. $O(2^K \cdot N^K)$
8. В 2 раза

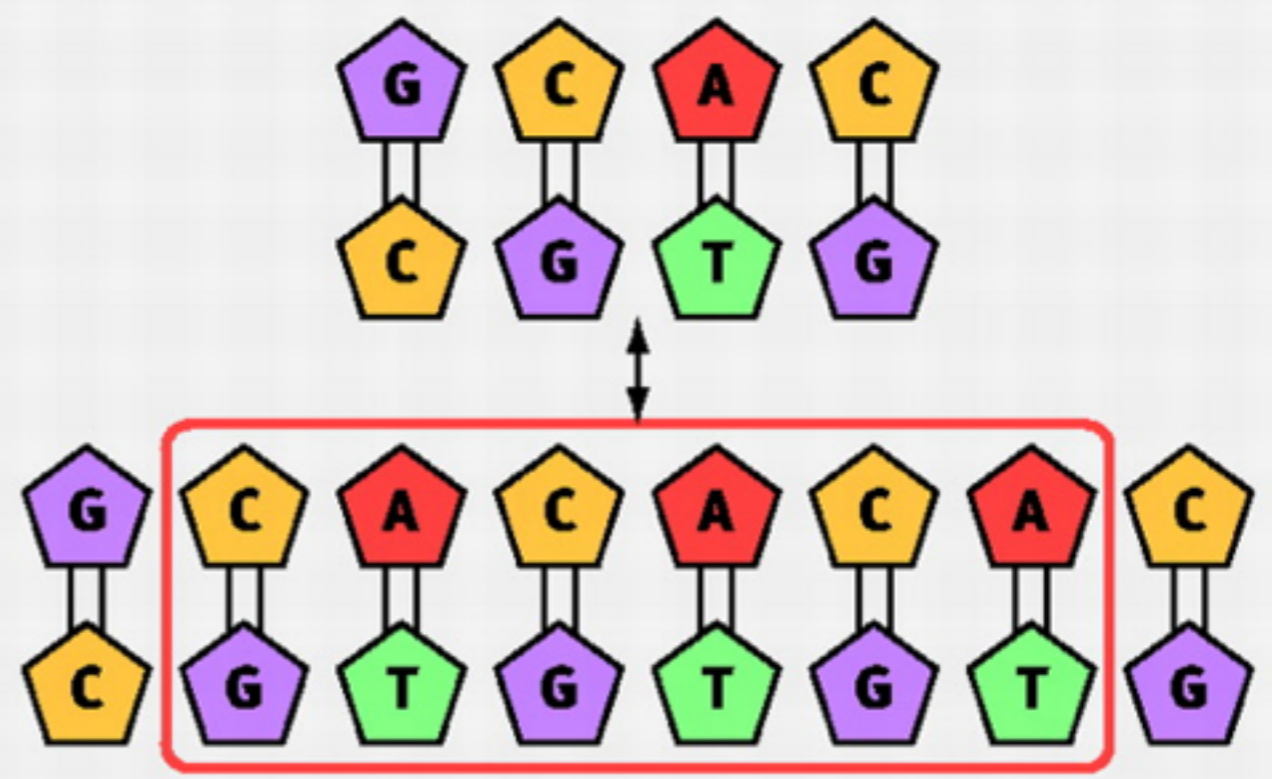
Single nucleotide polymorphism (SNP)



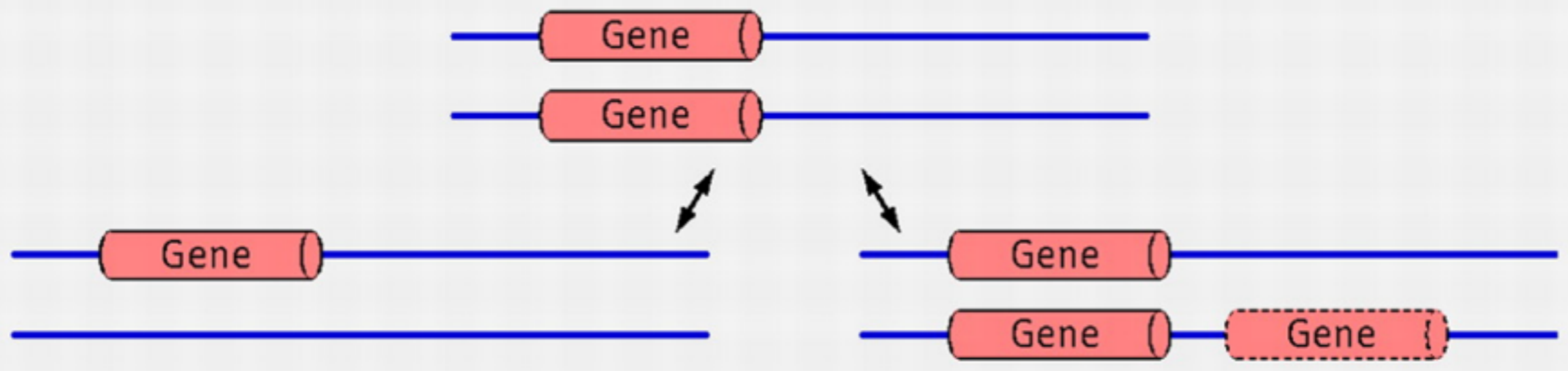
Insertion and deletion polymorphism (indel)



Nucleotide repeat polymorphism



Copy number variation



Deletion

Duplication

Геномные перестройки

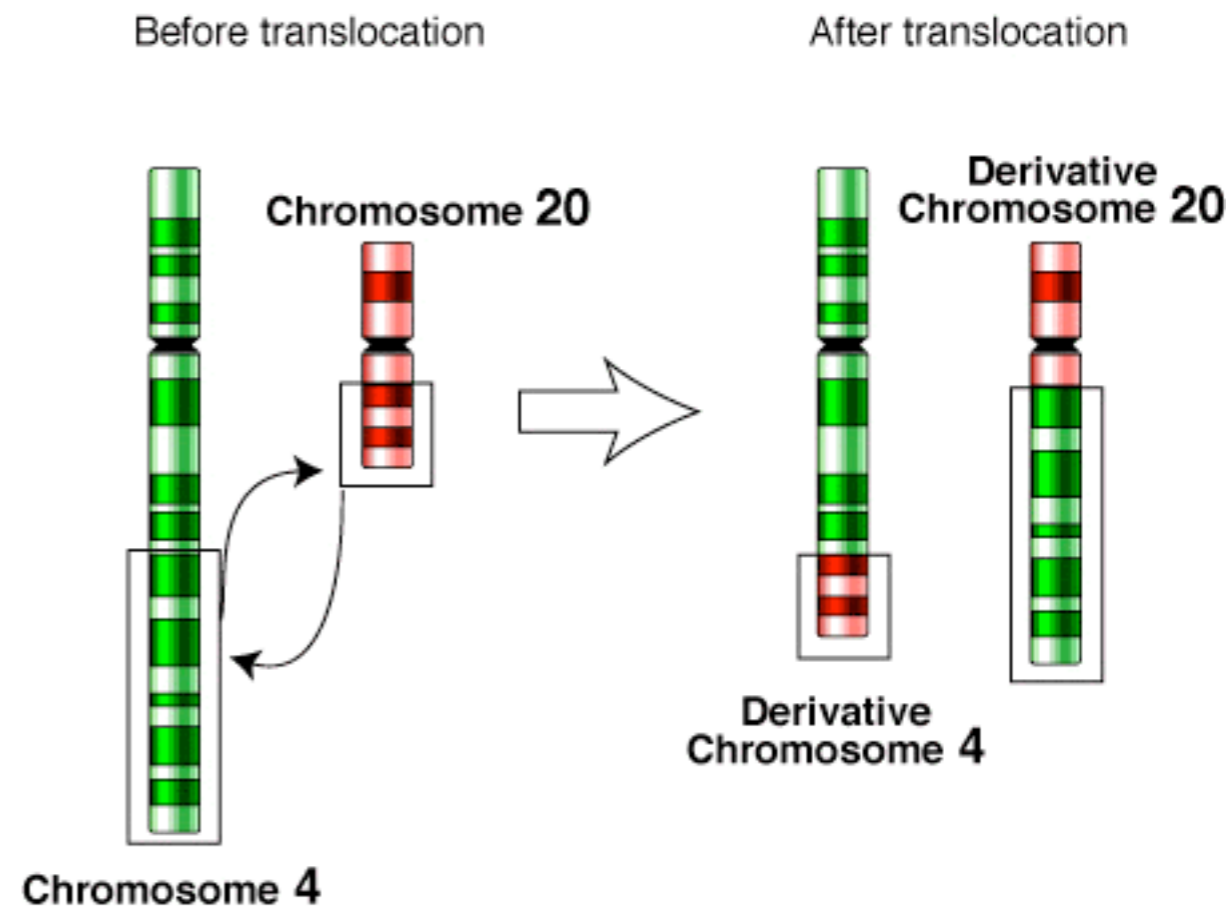
Genome Rearrangements:

реверсия (инверсия)

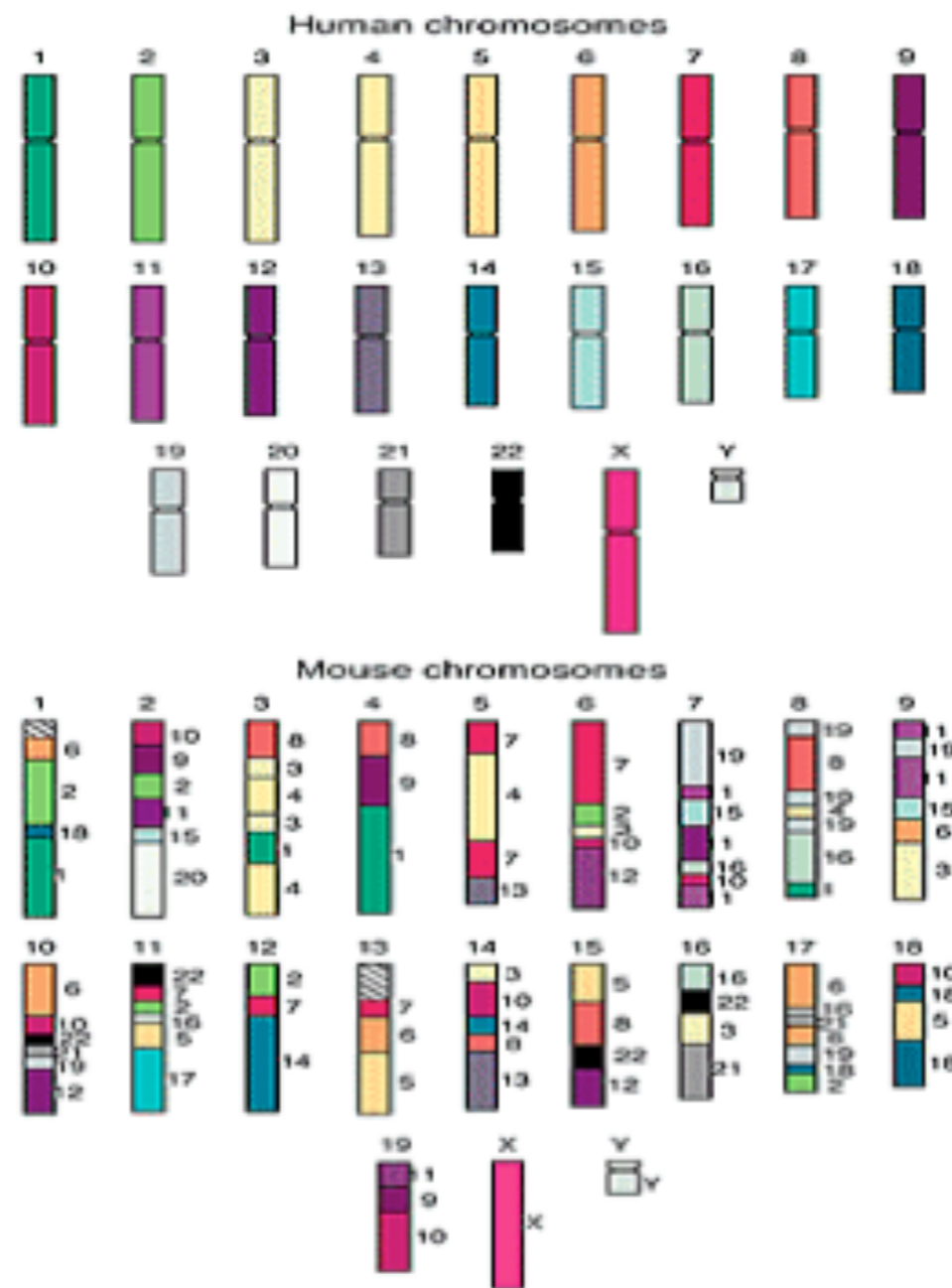
транслокация

слияние

расщепление

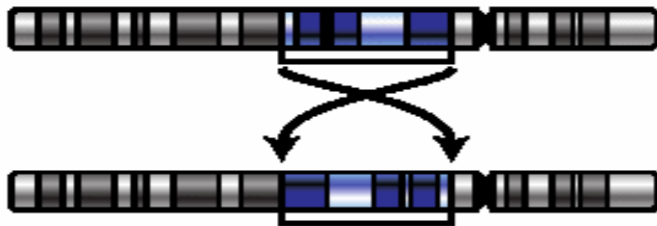


Геномные перестройки



Реверсия

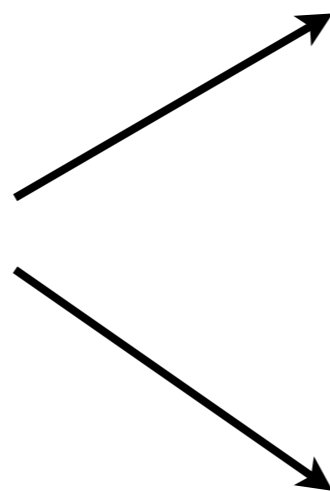
5' АТГ**ССТГТА**СТА 3'
3' ТА**СГГА**САТГАТ 5'



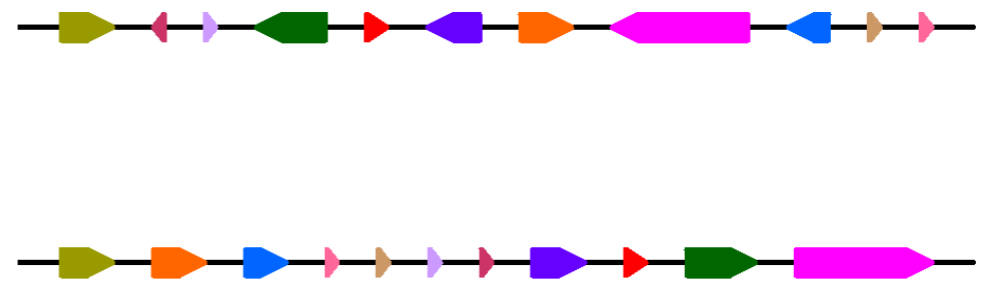
5' АТГ**ТАСА**ГГСТА 3'
3' ТА**САТГ**ТССГАТ 5'

Эволюционный сценарий

Неизвестный
предок
80 млн. лет
назад



X-хромосома мыши

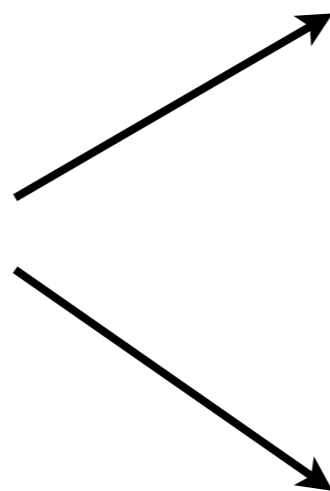


X-хромосома человека

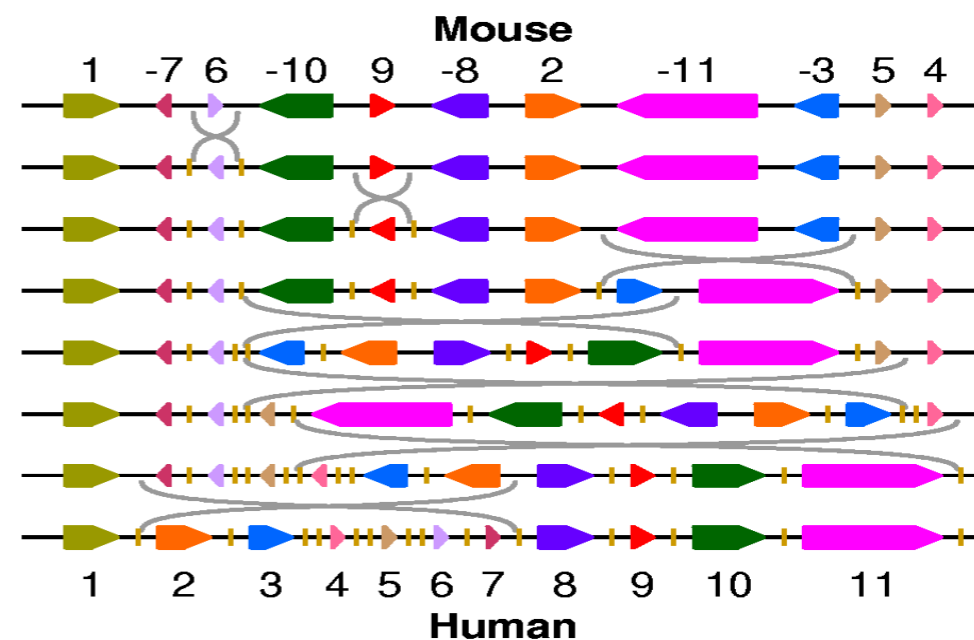


Эволюционный сценарий

Неизвестный
предок
80 млн. лет
назад

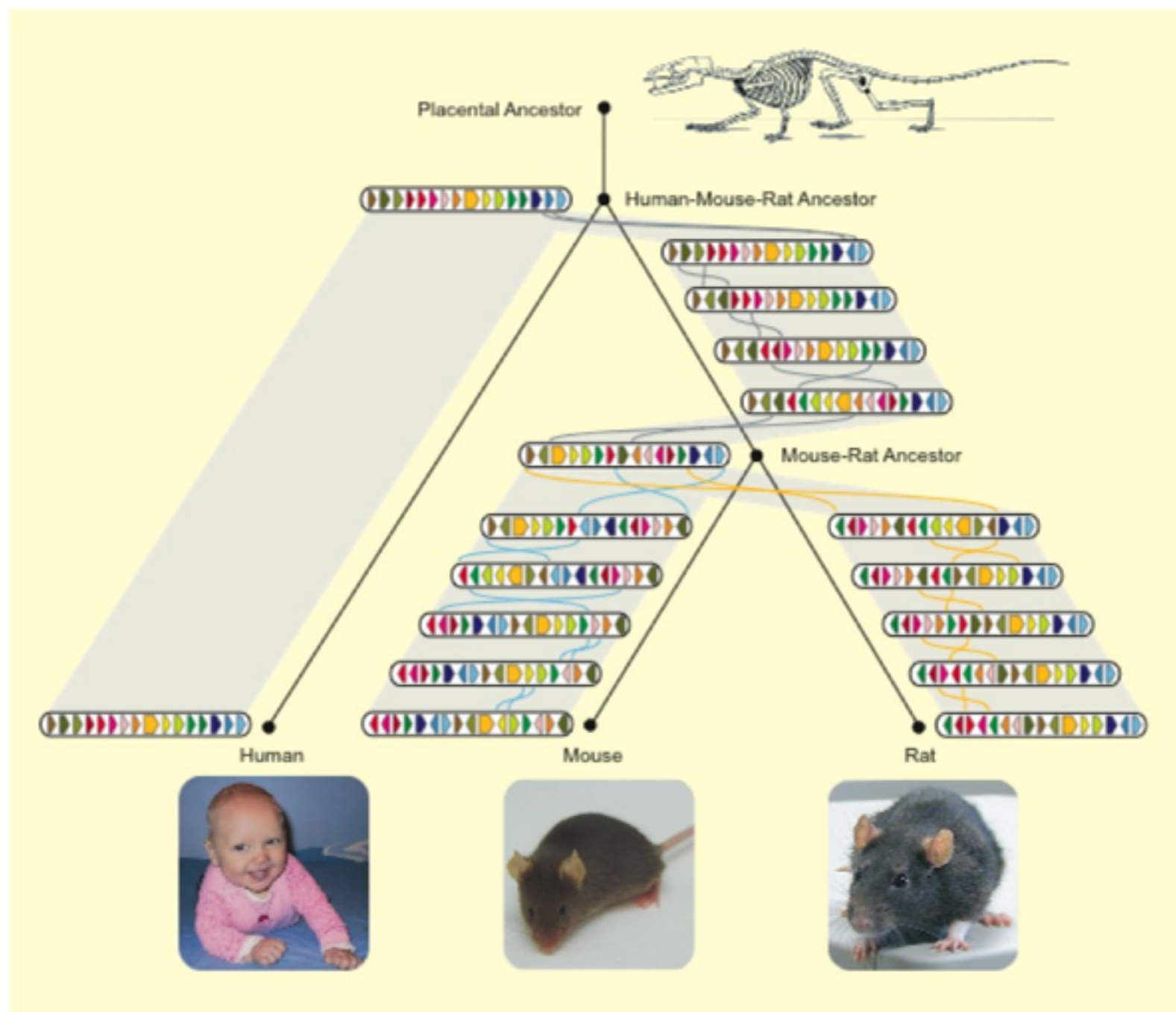


X-хромосома мыши



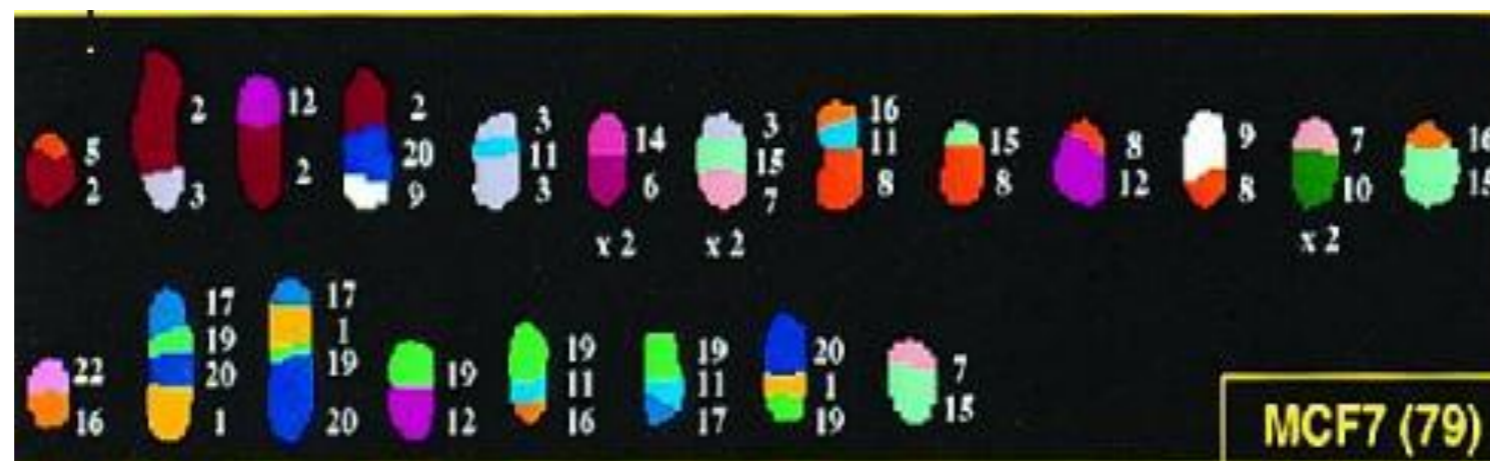
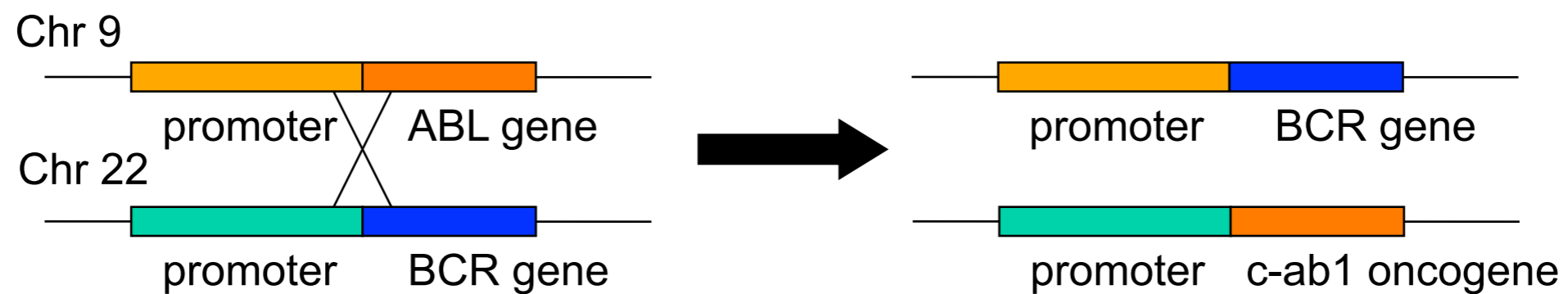
X-хромосома человека

Геномы предков



Точки перестройки

Филадельфийская хромосома



Сортировка реверсиями

1 8 9 3 2 7 6 5 4

1 2 3 9 8 7 6 5 4

1 2 3 4 5 6 7 8 9

Сортировка реверсиями

1 8 9 3 2 7 6 5 4

1 2 3 9 8 7 6 5 4

1 2 3 4 5 6 7 8 9

NP-hard

Знаковая сортировка реверсиями

2	<u>-4</u>	<u>-3</u>	5	-8	-7	-6	1
<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	-8	-7	-6	1
-5	-4	-3	-2	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>1</u>
<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>	6	7	8
1	2	3	4	5	6	7	8

Знаковая сортировка реверсиями

2	<u>-4</u>	<u>-3</u>	5	-8	-7	-6	1
<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	-8	-7	-6	1
-5	-4	-3	-2	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>1</u>
<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>	6	7	8
1	2	3	4	5	6	7	8

Reversal distance — минимальное количество реверсий, необходимых для преобразования одного генома в другой.

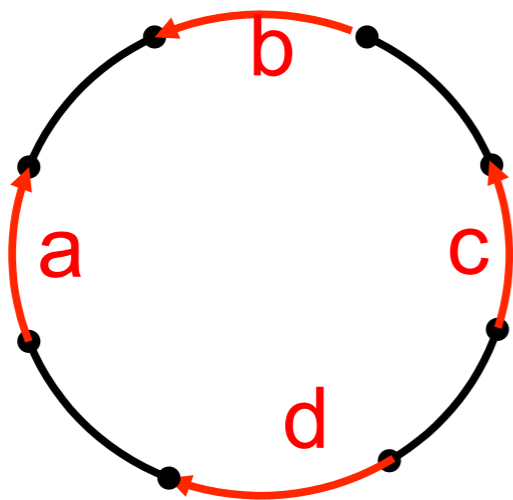
Знаковая сортировка реверсиями

2	<u>-4</u>	<u>-3</u>	5	-8	-7	-6	1
<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	-8	-7	-6	1
-5	-4	-3	-2	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>1</u>
<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>	6	7	8
1	2	3	4	5	6	7	8

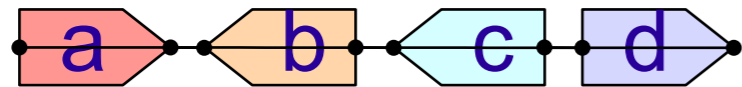
Reversal distance — минимальное количество реверсий, необходимых для преобразования одного генома в другой.

Упрощение

Хромосома — цикл.

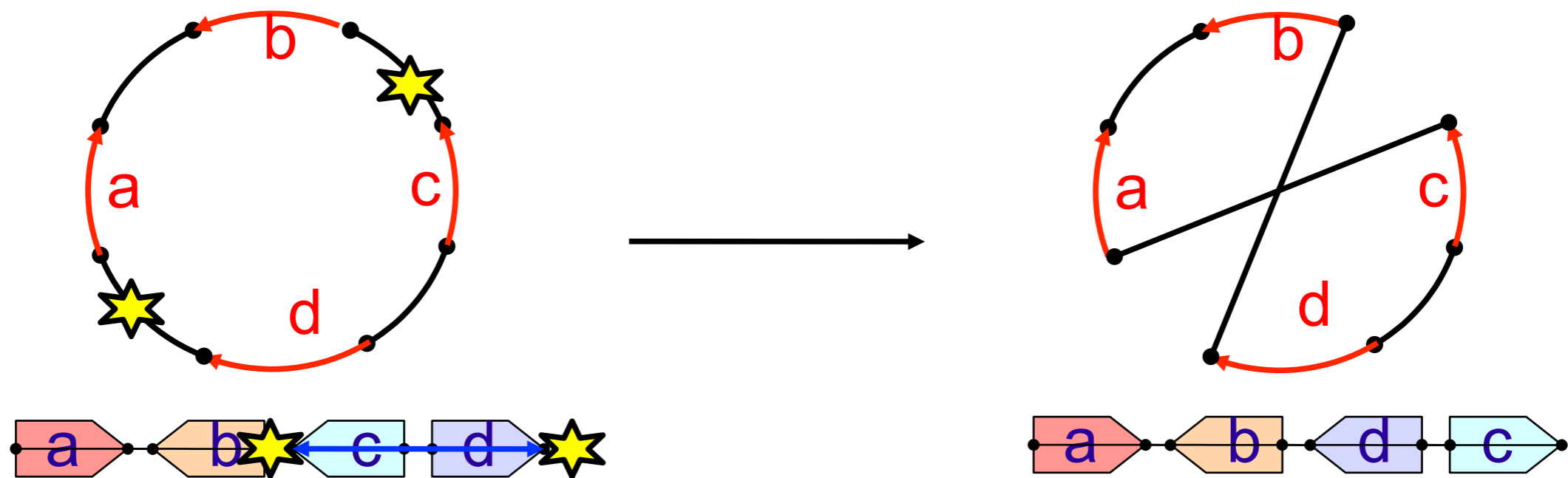


Направленные **красные** рёбра представляют блоки (гены).



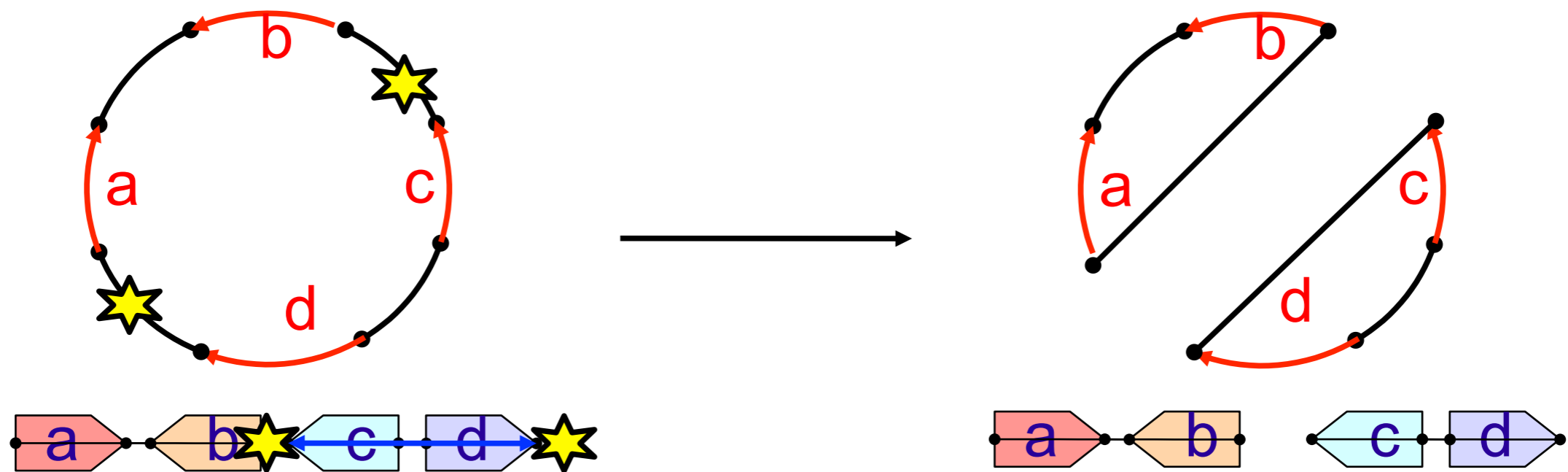
Ненаправленные **чёрные** рёбра соединяют соседние блоки.

Реверсия



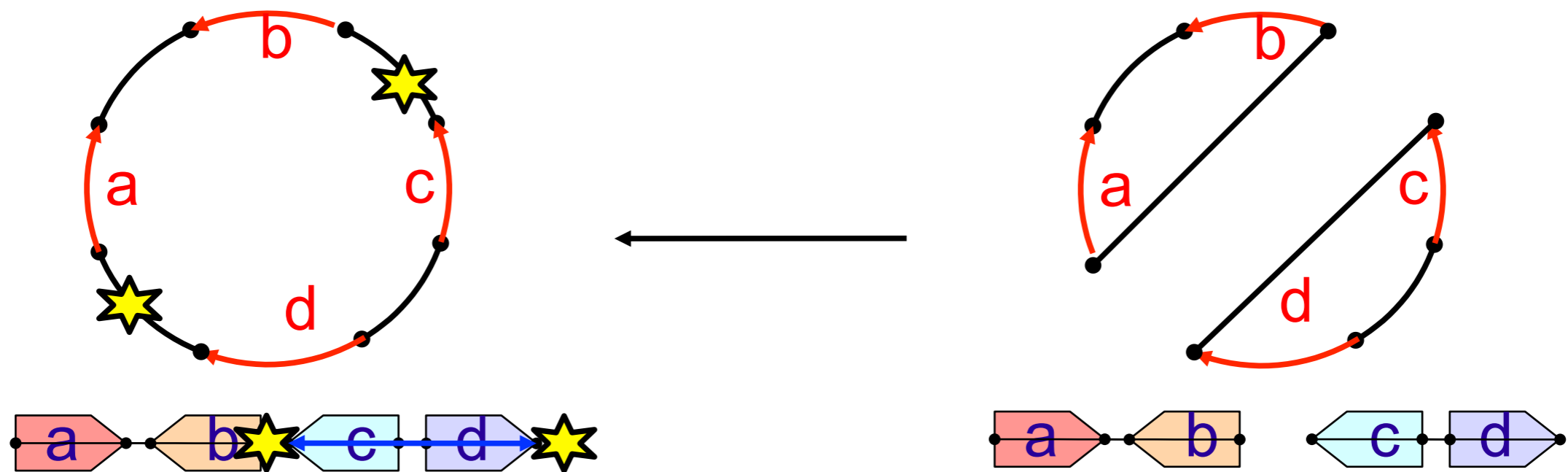
Реверсия заменяет два чёрных рёбра.

Расщепление



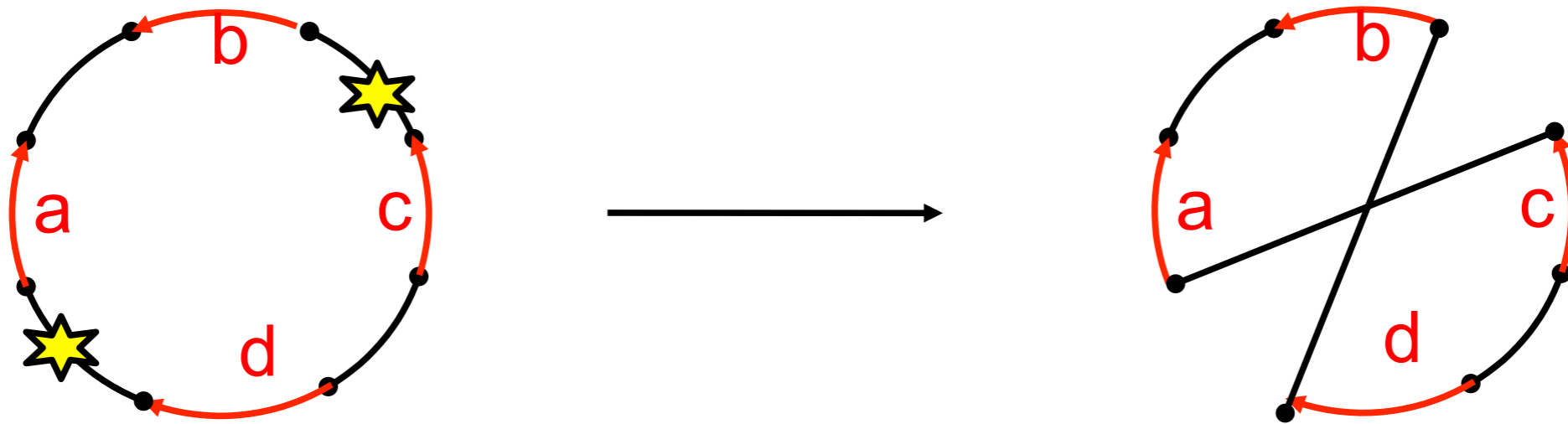
Расщепление заменяет два чёрных рёбра
и разделяет один цикл на два.

Слияние



Слияние заменяет два чёрных рёбра
и соединяет два цикла в один.

2-Breaks



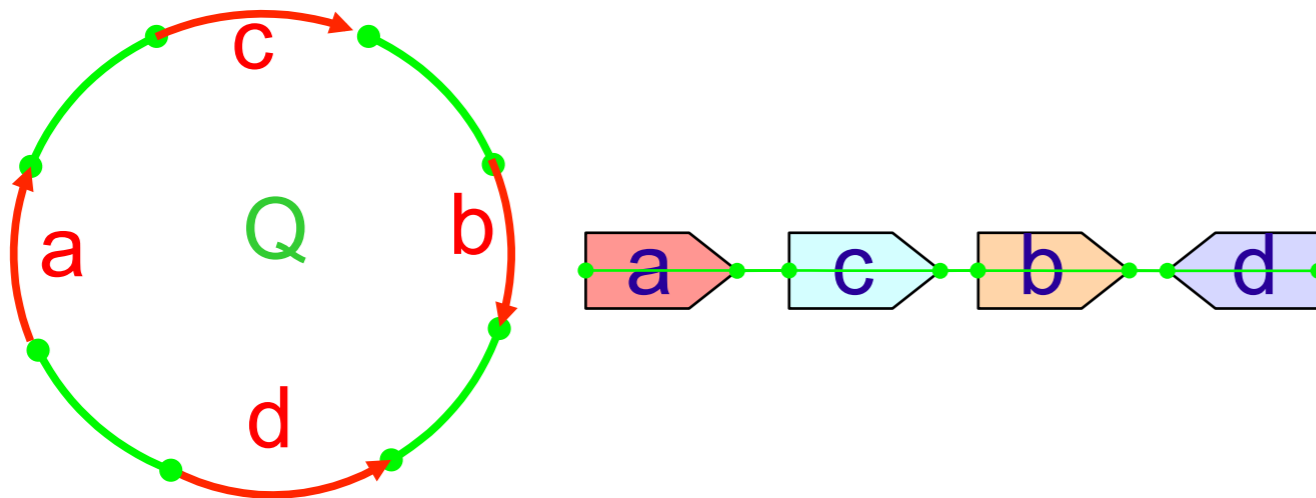
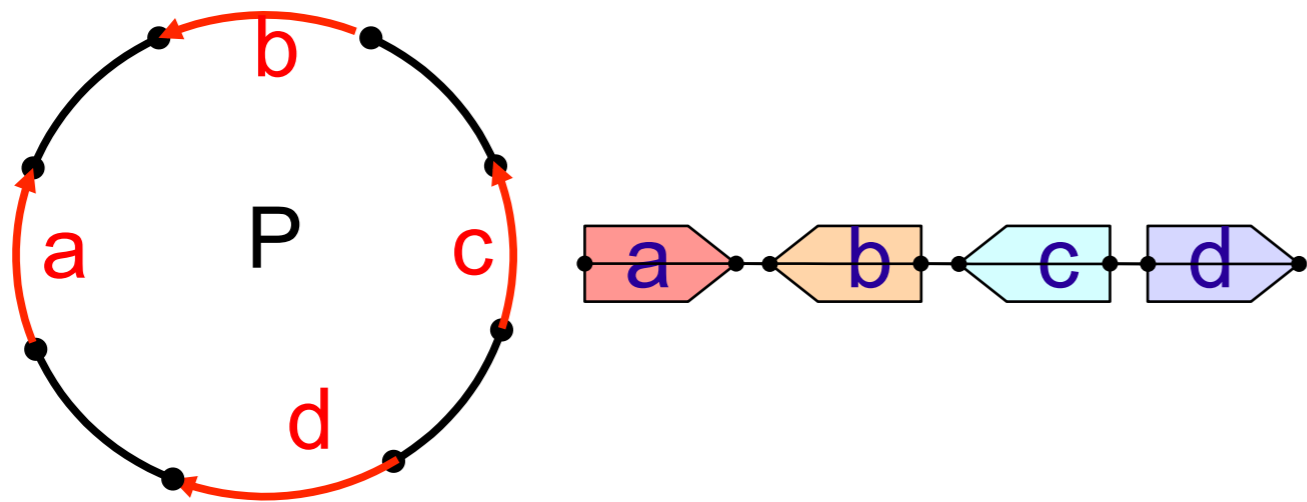
2-Break (double cut-and-join, DCJ) — замена 2-х любых чёрных ребер на 2 других чёрных ребра, соединяющие те же 4 вершины.

Реверсии, транслокации, слияния и расщепления — все возможные 2-breaks.

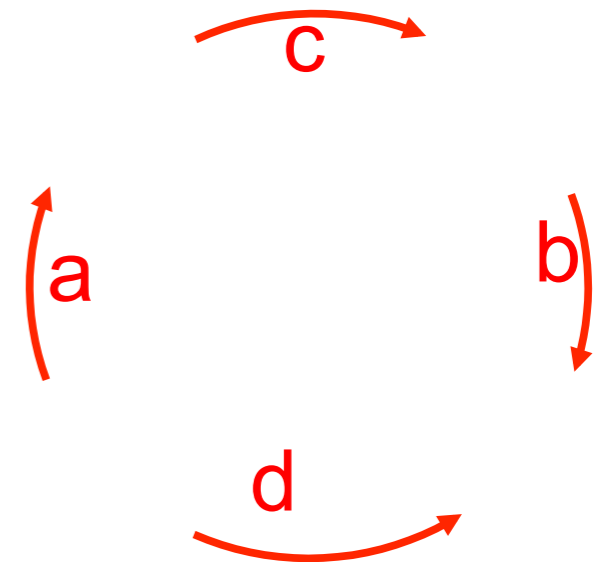
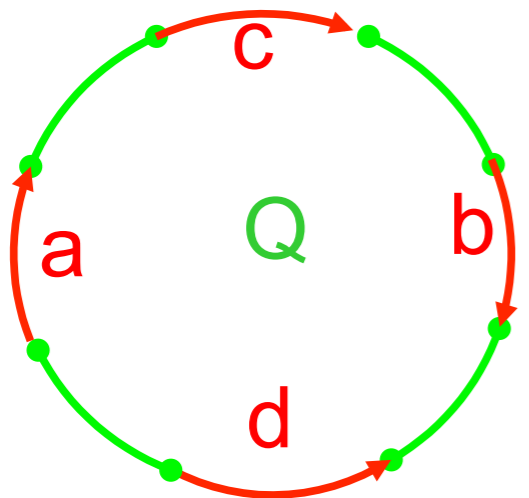
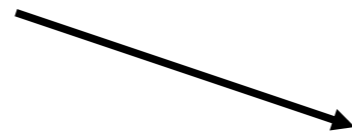
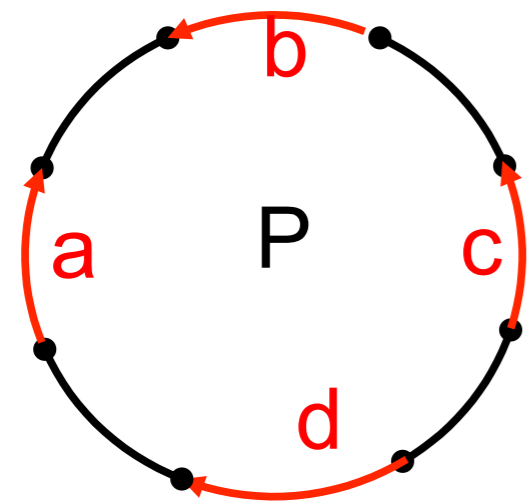
2-Break Distance

Минимальное количество 2-breaks,
необходимых для преобразования одного
генома в другой.

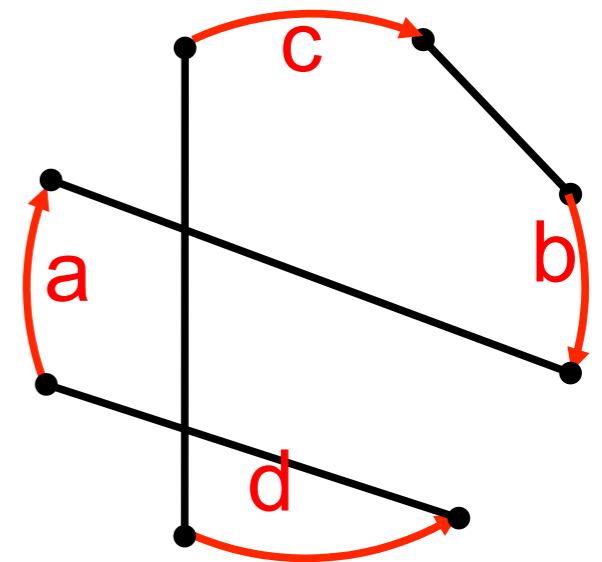
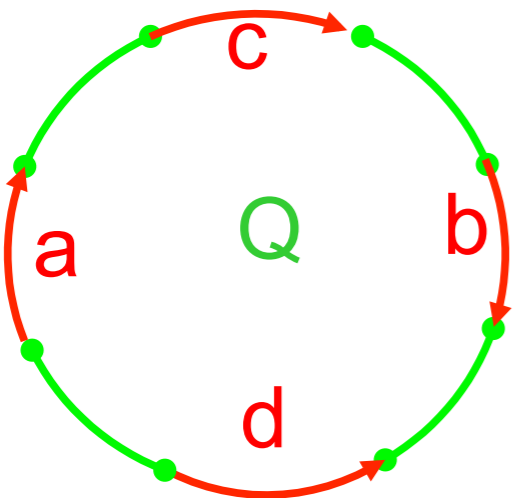
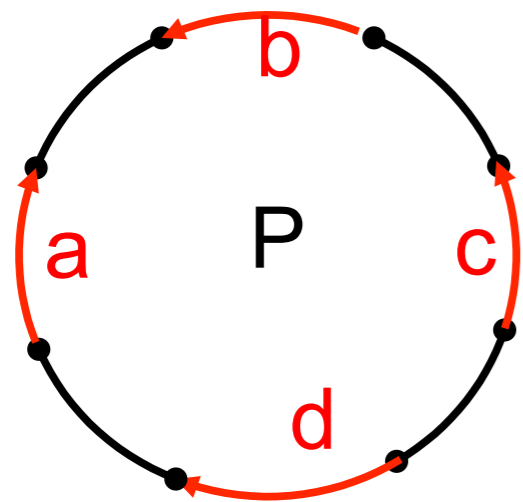
2-Break Distance



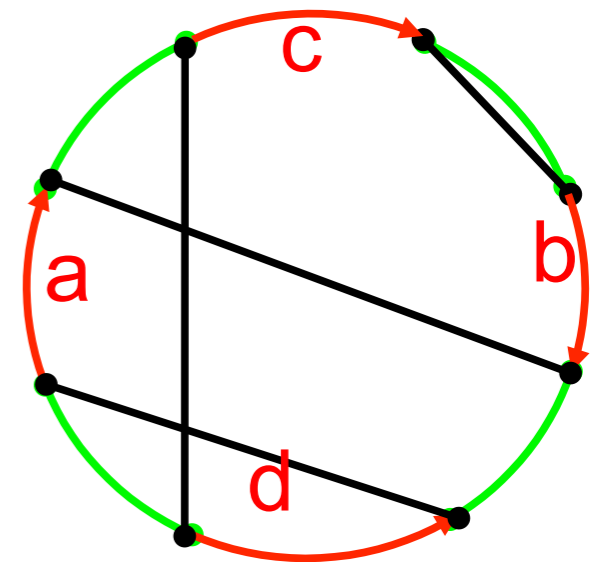
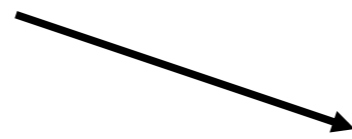
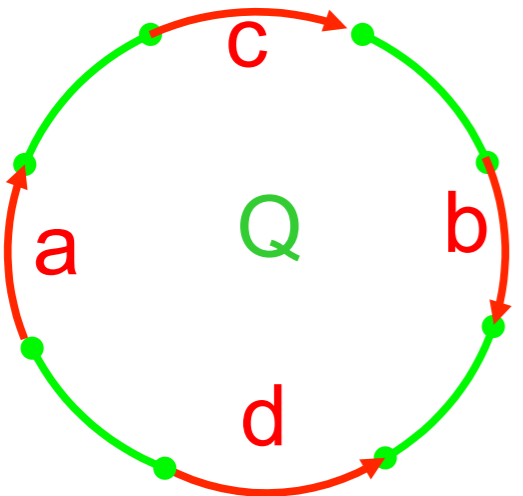
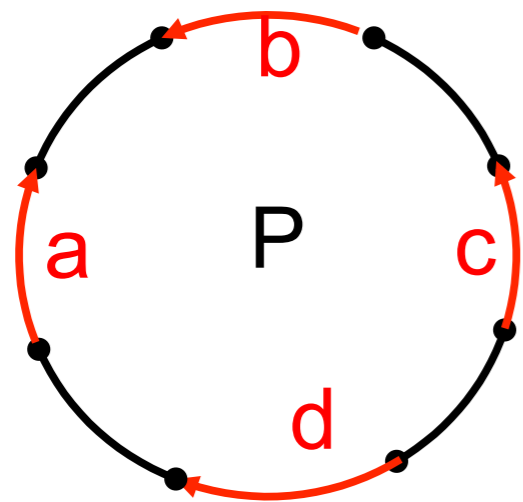
2-Break Distance



2-Break Distance



2-Break Distance

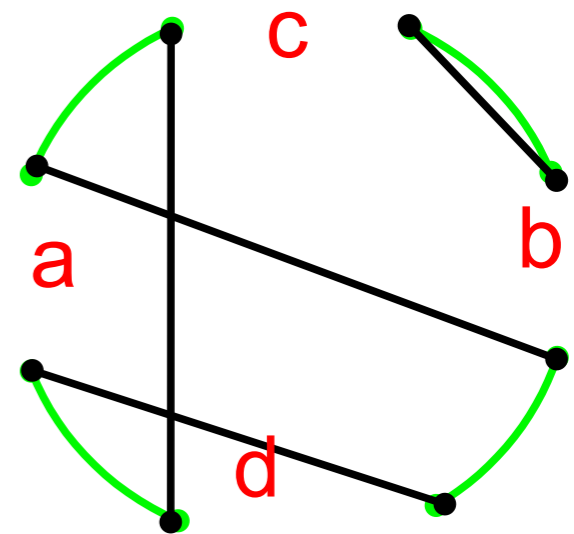


Breakpoint Graph

Breakpoint Graph

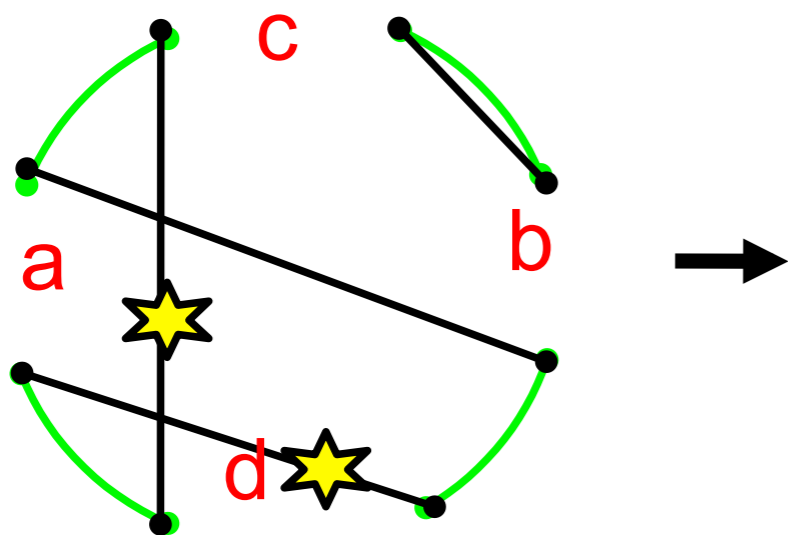
Рёбра образуют совершенные паросочетания (рёберные покрытия).

Рёбра формируют чёрно-зелёные альтернирующие ЦИКЛЫ.



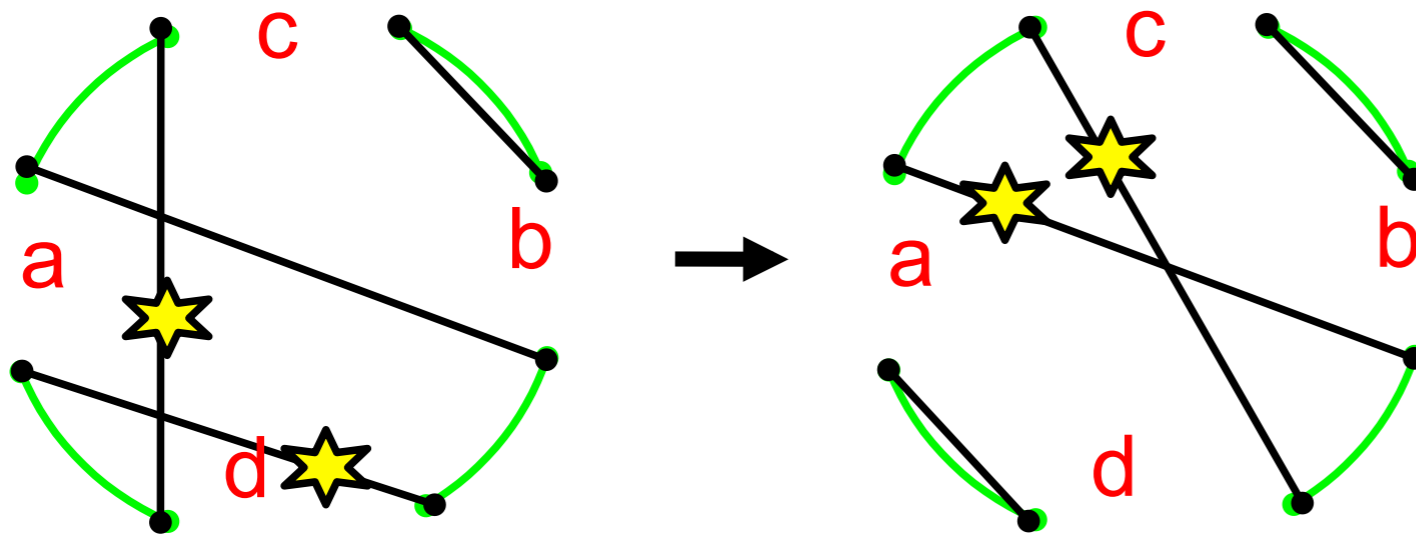
$$\text{cycles}(P, Q) = 2$$

Перестройки



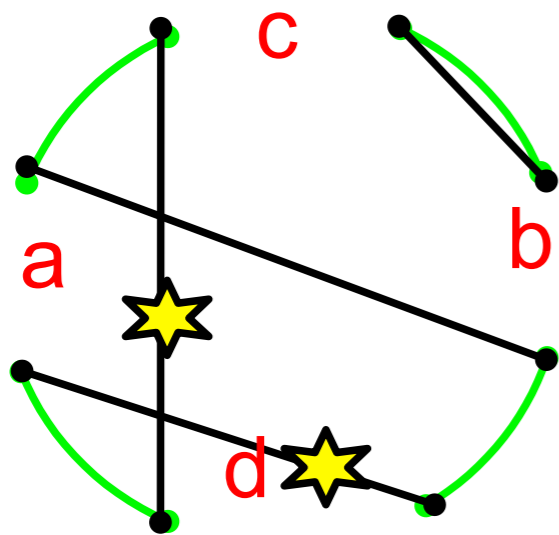
2 цикла

Перестройки

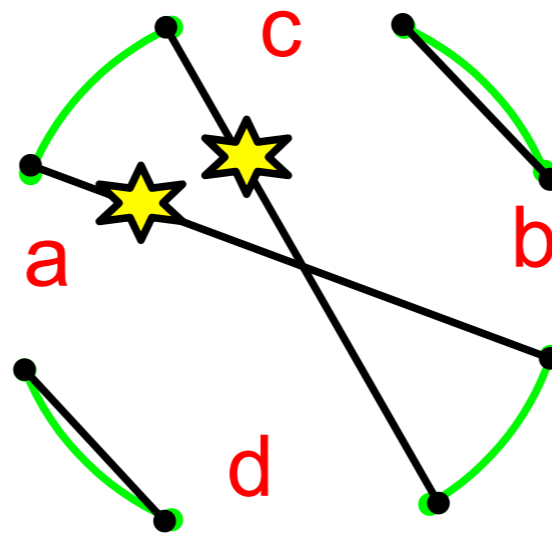


2 цикла

Перестройки

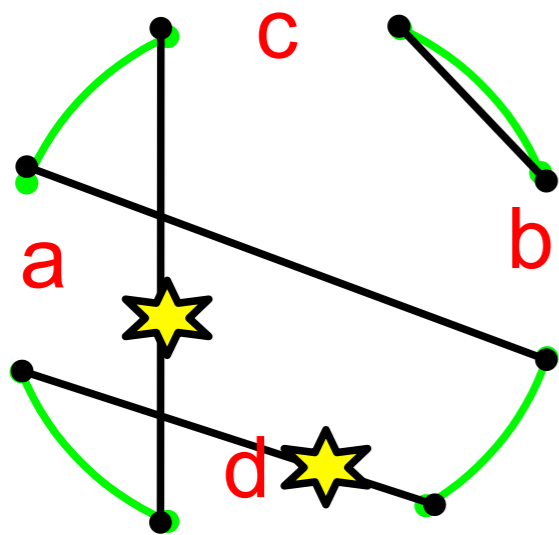


2 цикла

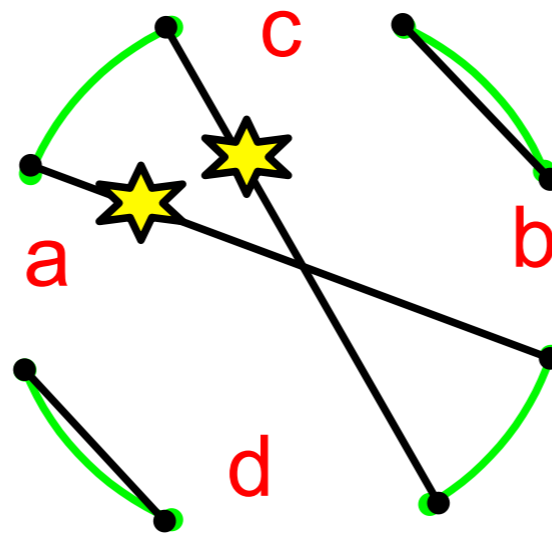


3 цикла

Перестройки



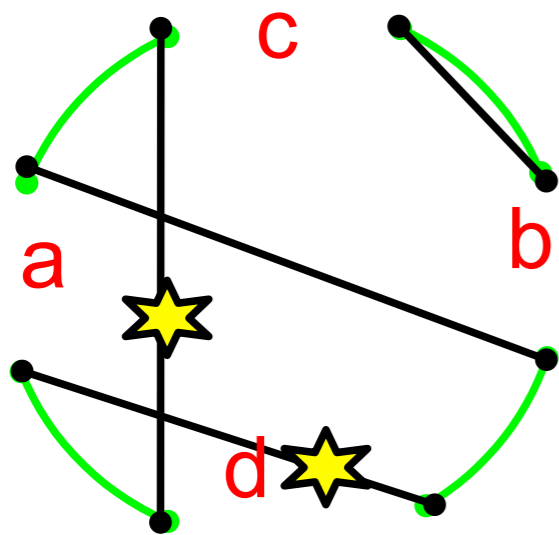
2 цикла



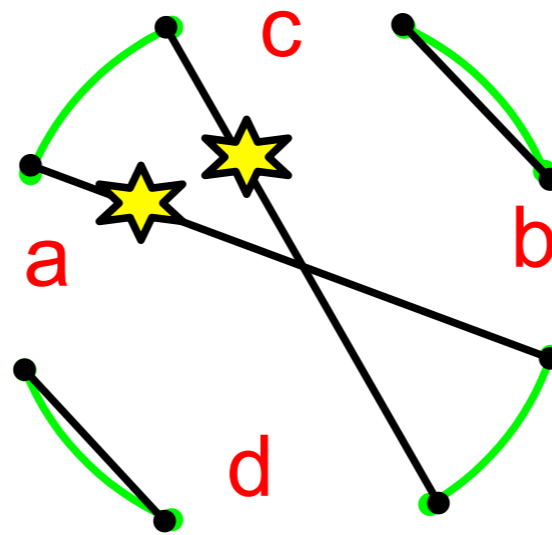
3 цикла



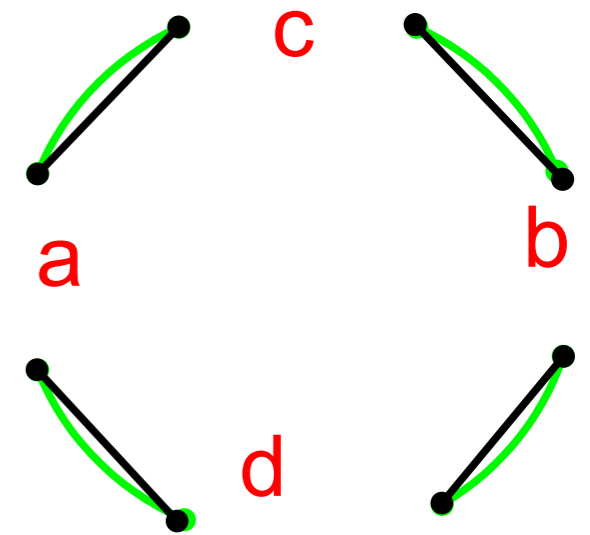
Перестройки



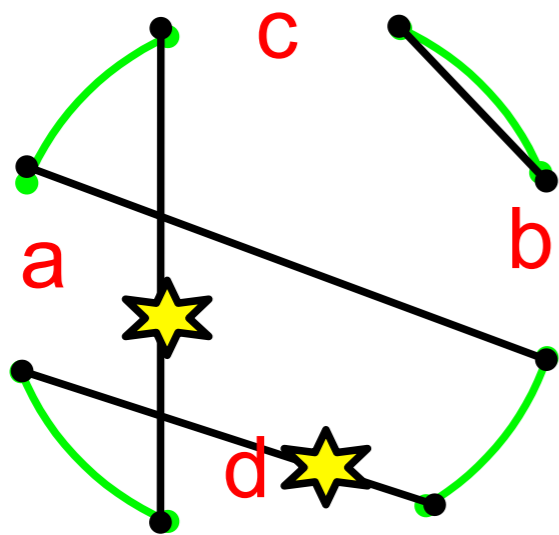
2 цикла



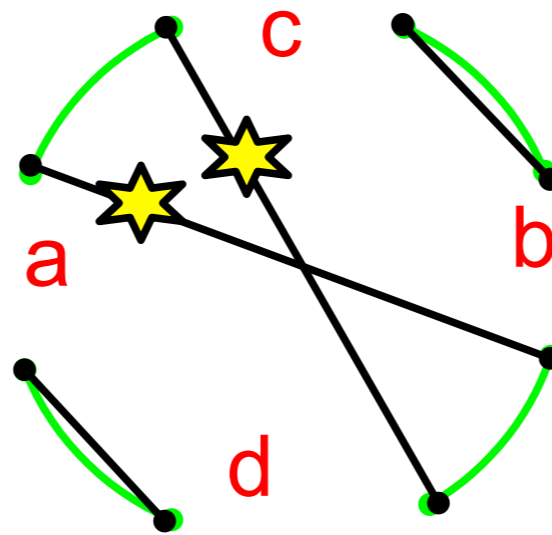
3 цикла



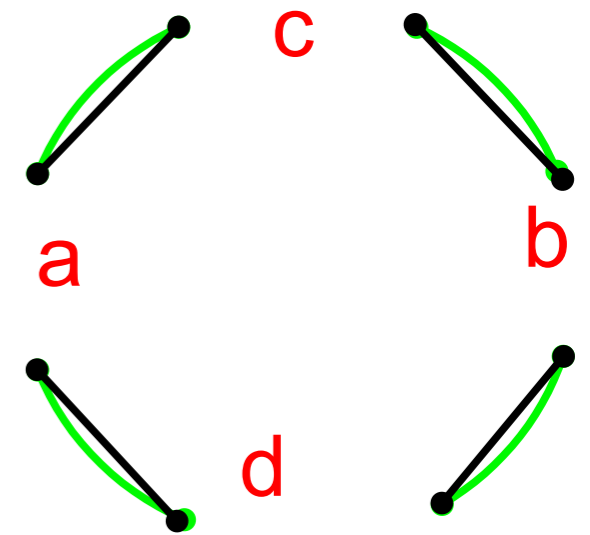
Перестройки



2 цикла

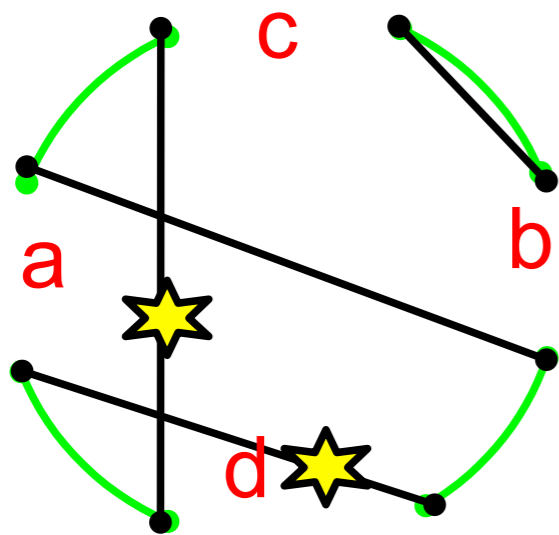


3 цикла

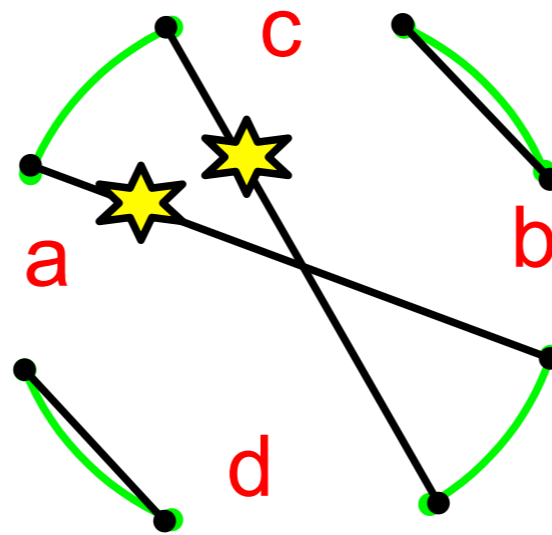


4 цикла

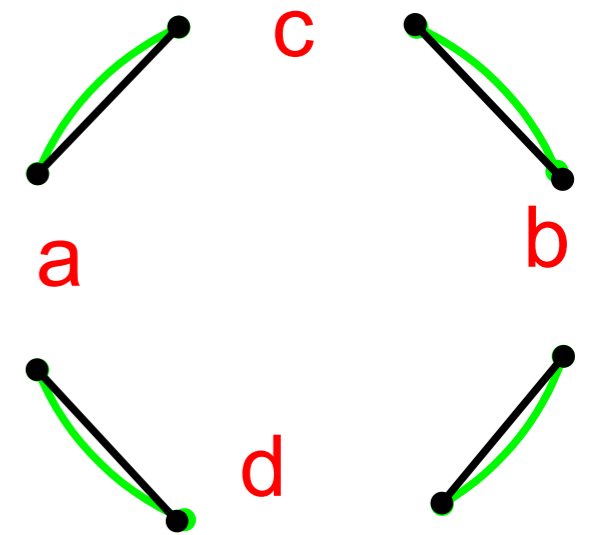
Перестройки



2 цикла



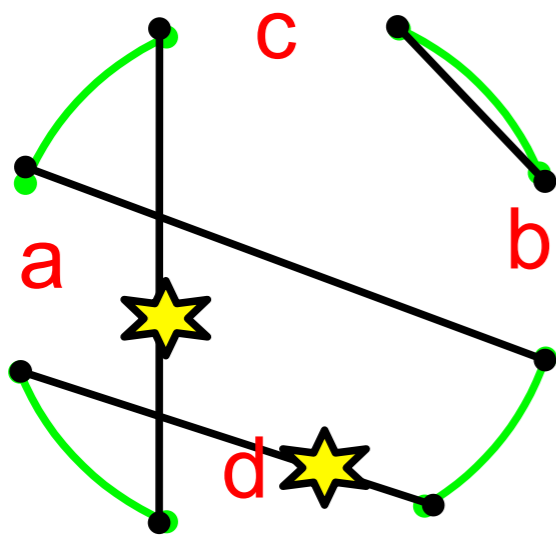
3 цикла



4 цикла

= #blocks

Каждый 2-break



Добавляет 2 чёрных ребра \Rightarrow
создаёт максимум 2 новых цикла.

Удаляет 2 чёрных ребра \Rightarrow
удаляет хотя бы 1 старый цикл.

Расстояние 2-break

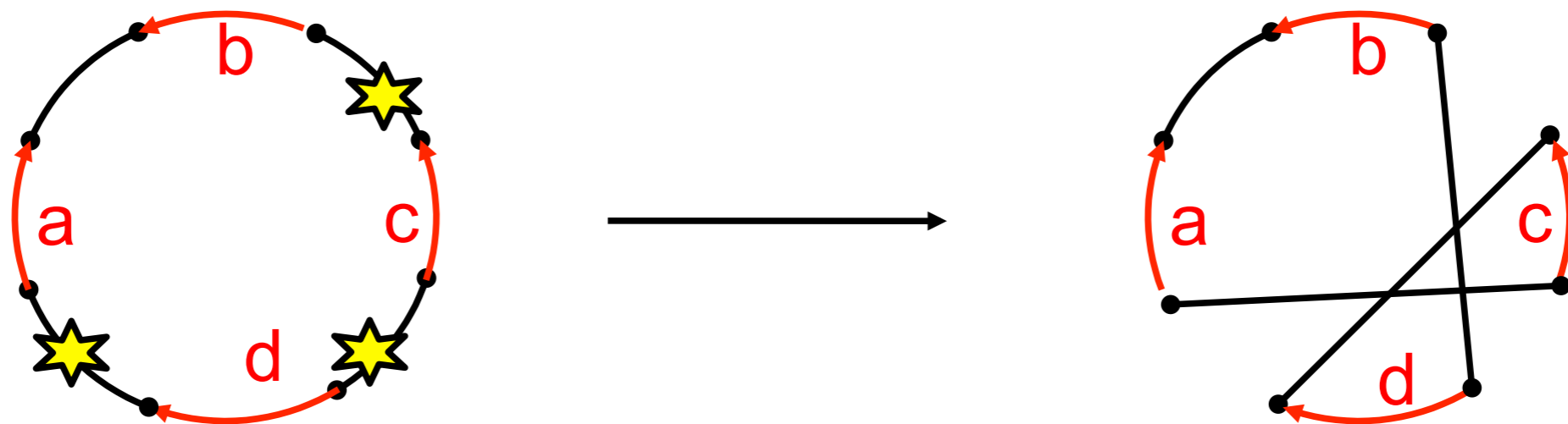
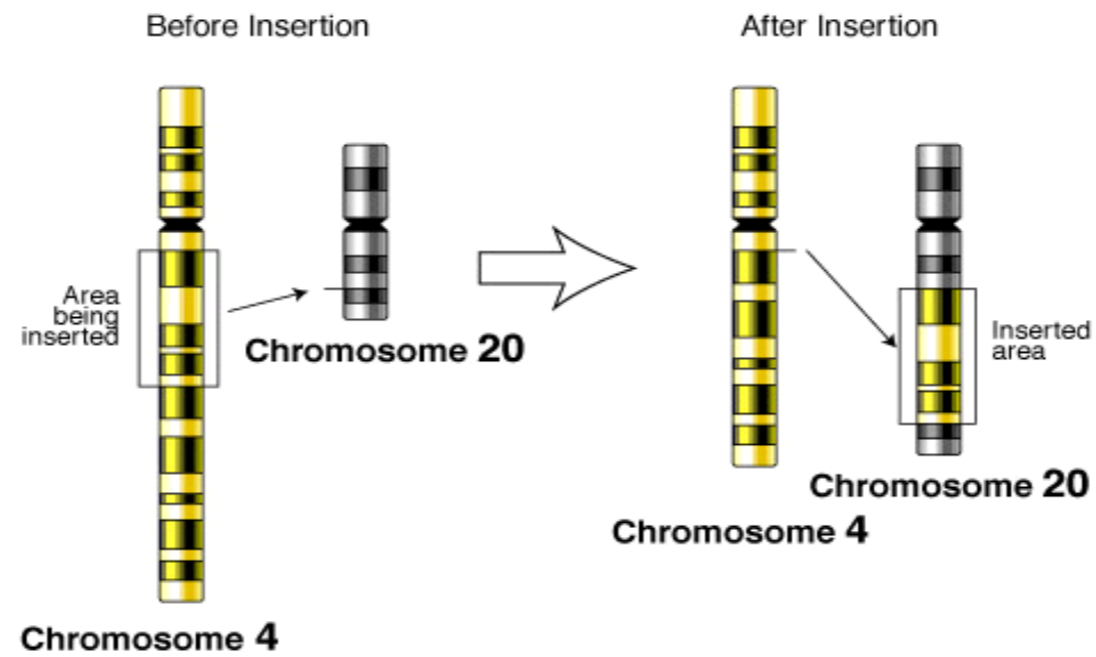
Каждый 2-break увеличивает количество циклов хотя бы на 1.

Каждый нетривиальный цикл может быть разбит на два цикла при помощи 2-break.

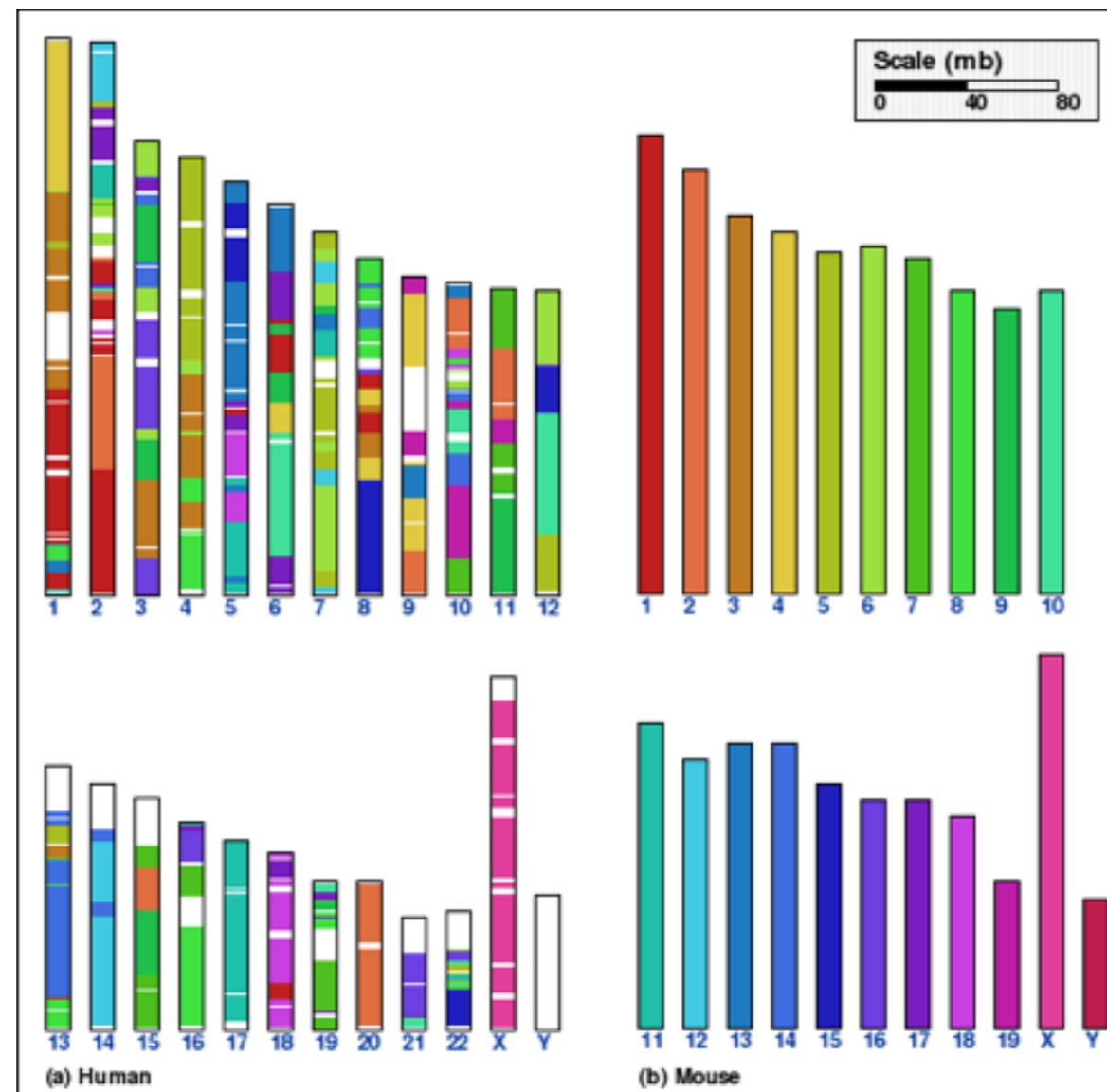
Для преобразования P в Q необходимо увеличить количество циклов на $\#blocks - cycles(P, Q)$.

$$d_2(P, Q) = \#blocks - cycles(P, Q)$$

Транспозиции

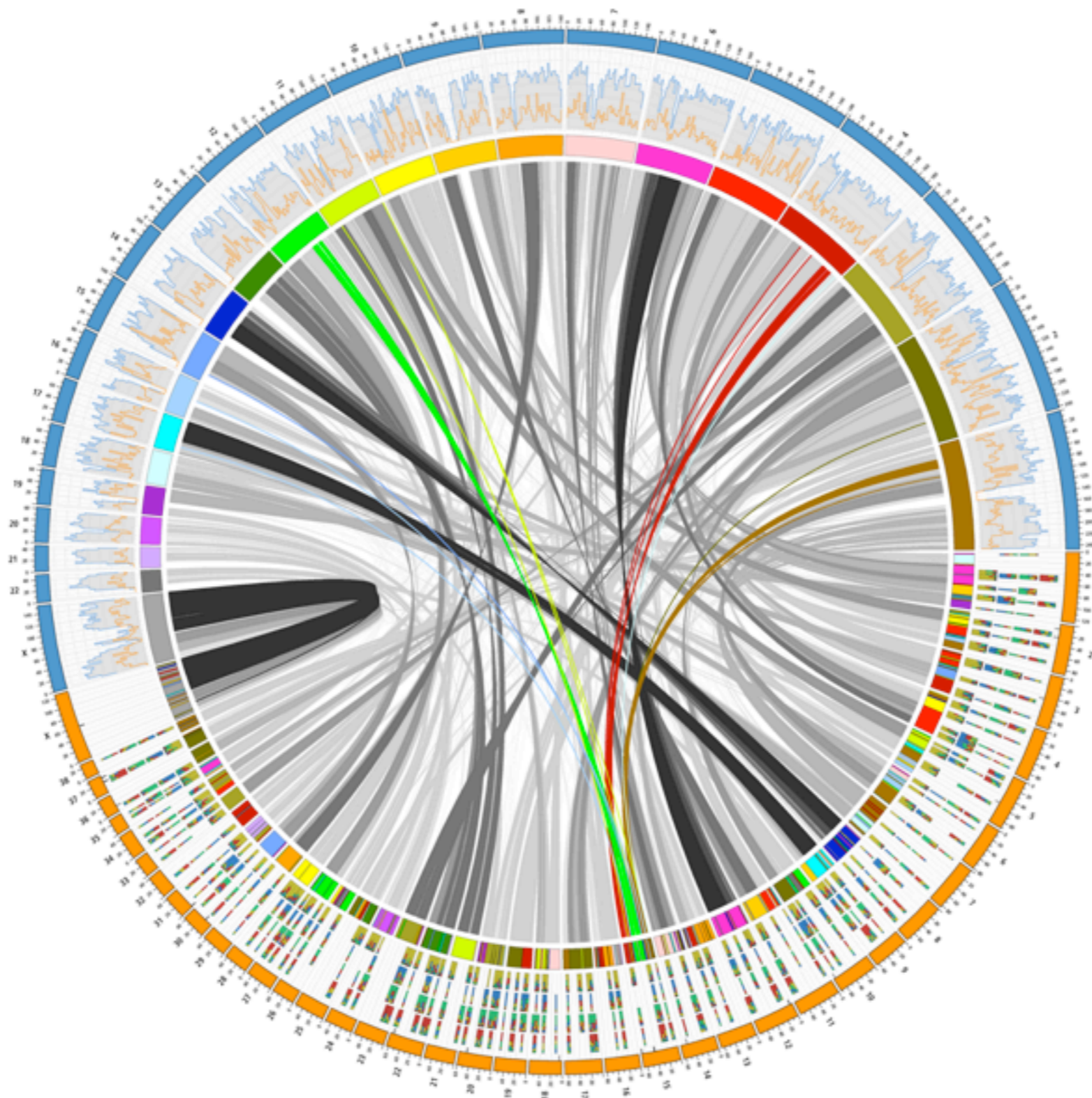


Synteny Blocks



Синтения – это структурное сходство групп сцепления генов у организмов разных биологических видов.

Сравнительная геномика



Формат обучения

12 лекций по воскресеньям

Квизы для самопроверки

Домашние задания и вопросы онлайн

<http://rosalind.info/classes/enroll/ff45302de4/>

